Available online at www.sciencedirect.com

ScienceDirect

Current Opinion in Structural Biology

New opportunities in integrative structural modeling

Daniel S. Ziemianowicz[1,2] and Jan Kosinski[1,2,3]

Abstract

Integrative structural modeling enables structure determination of macromolecules and their complexes by integrating data from multiple sources. It has been successfully used to characterize macromolecular structures when a single structural biology technique was insufficient. Recent developments in cellular structural biology, including in-cell cryo-electron tomography and artificial intelligence-based structure prediction, have created new opportunities for integrative structural modeling. Here, we will review these opportunities along with the latest developments in integrative modeling methods and their applications. We also highlight open challenges and directions for further development.

Addresses

[1] European Molecular Biology Laboratory Hamburg, Notkestraße 85, 22607, Hamburg, Germany

[2] Centre for Structural Systems Biology (CSSB), Notkestraße 85, 22607, Hamburg, Germany

[3] Structural and Computational Biology Unit, European Molecular Biology Laboratory, Meyerhofstraße 1, 69117, Heidelberg, Germany

Corresponding author: Kosinski, JanEuropean Molecular Biology Laboratory Hamburg, Notkestraße 85, 22607, Hamburg, Germany. (jan.kosinski@embl.de)

(Ziemianowicz D.S.)

(Kosinski J.)

Current Opinion in Structural Biology 2022, **77**:102488

This review comes from a themed issue on **Theory and Simulation/ Computational Methods (2023)**

Edited by **Turkan Haliloglu** and **Gregory A. Voth**

For complete overview of the section, please refer the article collection - Theory and Simulation/Computational Methods (2023)

Available online 21 October 2022

https://doi.org/10.1016/j.sbi.2022.102488

0959-440X/© 2022 The Author(s). Published by Elsevier Ltd. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).

Introduction

Integrative modeling allows macromolecular structure determination by combining data from multiple experimental and computational techniques [1]. In essence, empirical data guides the assembly of a protein complex from its component structures. For example, atomic structures may be fit into the 3D density obtained by cryo-electron microscopy (cryo-EM) or scattering profiles by small-angle X-ray scattering (SAXS) and simultaneously orientated with distance restraints obtained by cross-linking mass spectrometry (XL-MS) (Figure 1). Depending on data quality, integrative models can inform the general architecture, find evolutionary relationships, localize active sites and near-atomic interactions, and generate hypotheses on the mechanism(s) of action.

Integrative modeling has resolved the structure of many complexes (Figure 2). A prime example is the nuclear pore complex, first published in 2007 [2]. Since then, nuclear pore complexes from various species have been modeled with increasingly precise data, culminating with near-atomic yeast and human models [3–5]. Other recent examples include the Fanconi anemia core complex [6], yeast Smc5/6 holo-complex [7], and mycobacterial type VII secretion system [8] built using cryo-EM and XL-MS data; a ribozyme using NMR and SAXS data [9]; RAGE complexed with S100B using XL-MS, hydrogen-deuterium exchange, and mass spectrometry data [10]; and histone H3-H4 and RNA polymerase subcomplexes using restraints from a novel genetic interaction mapping technique [11].

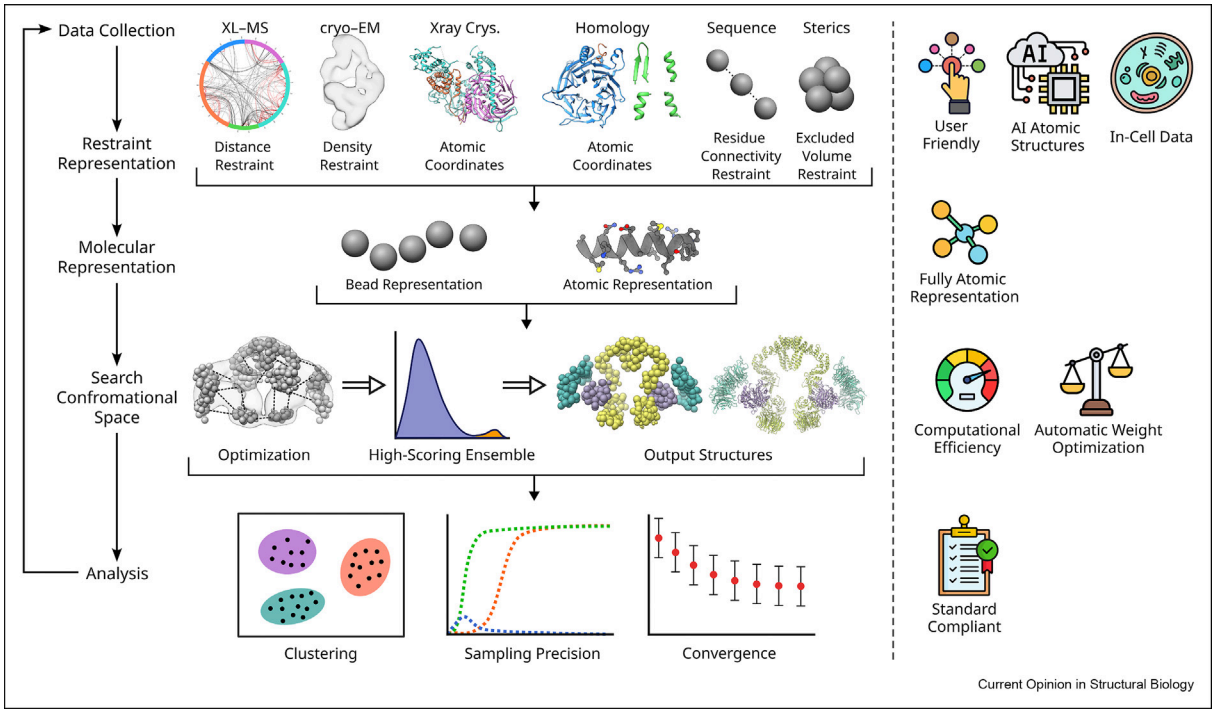
Many integrative modeling software tools are available, with Integrative Modeling Platform (IMP) [12] and HADDOCK [13] being the most popular. Derivative programs targeting different modeling challenges and methodological aspects include Python modeling interface [14] and our Assembline protocol [15], both based on IMP; M3 [16] which is based on HADDOCK; or IMProv [17], a graphical interface to IMP.

Here, we will review recent innovations and new opportunities arising from advances in experimental and computational techniques.

State-of-the-art in integrative structural modeling

The first step in integrative modeling, besides collecting experimental data, is choosing how to represent macromolecules during modeling (Figure 1). While small complexes can be represented at the atomic level, larger complexes must be represented more coarsely—such as spherical beads approximating individual or groups of residues—for computational efficiency. Multiple levels can be used simultaneously in a so-called multi-scale representation, with fine-grain details retained only for calculations that require accuracy [1,18]. IMP, PMI, and

Figure 1



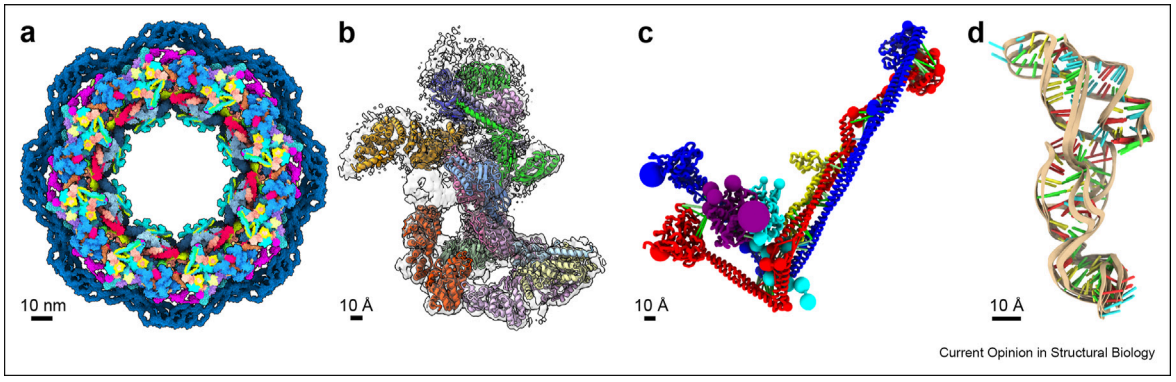
Schematic of a typical integrative modeling workflow (left) with opportunities and key areas for development (right).

Assembline use multi-scale representation to efficiently generate large models with atomic features. Though the choice of representation is usually dictated by the software and the modeler's intuition, an automatic optimization protocol based on local information density has been proposed [19].

The second step is representing experimental data as restraints—mathematical formulations that evaluate the fit of sampled models to the input data. Importantly, the restraints must account for uncertainty in the data, such as EM map resolution or the expected cross-link distance distribution, and be suitable for the scale of representation. Many experimental data types, such as EM, SAXS, or XL-MS, have already been implemented as restraints, with more sophisticated forms being developed [20,21].

Third, an optimization procedure searches the space of structural configurations for models that best fit the data, based on a scalar scoring function constructed from

Figure 2



Examples of recent integrative models. **(a)** Human nuclear pore complex model (PDB ID: 7R5K). **(b)** Fanconi anemia core complex (PDB DEV ID: 00000055). **(c)** Smc5/6 complex showing multi-scale atomic and bead representation (PDB DEV ID: 00000081). Green bars indicate crosslinks used for modeling. **(d)** Neurospora Varkud satellite ribozyme (PDB DEV ID: 00000067).

the restraints. While optimization procedures have matured [1,13], computational efficiency is still lacking for large and fine-grained models. The scoring function is usually a linear combination of individual restraints [13,15] with manually selected weights. Methods based on Bayesian inference have been proposed [12–15,21] with weights optimized automatically based on the estimated quality of the data. However, only a few restraint types and software have Bayesian formulations, and the resulting function is often still a linear combination of Bayesian and conventional terms. Thus, an integrative model is not only a hypothesis over coordinates but also the relative weighing of data types. In our opinion, automated weighting is one of the biggest challenges in integrative modeling that requires further advancement.

Last, integrative models must be assessed for precision, consistency with the data, and exhaustiveness of conformational sampling. A formalized solution is included in the imp-sampcon toolkit [22], recently improved to determine local precision [23].

New opportunities from artificial intelligence-based structure prediction

Traditionally, atomic structures for integrative modeling came mainly from X-ray crystallography, NMR, and cryo-EM. If no experimental structures existed, homology modeling could be used to predict structures based on homologous proteins as templates. Yet, this was a relatively time-consuming process limited by the data available. If no atomic structures could be obtained, spherical beads would be used to approximate residues or entire subunits. As a result, integrative models were limited in precision, with some subunits retaining "bead" representation, or having uncertain localization and unresolved steric clashes. Although fully atomic integrative modeling programs exist, such as HADDOCK, they are limited by the quality of input structures and the accuracy of the optimization algorithms.

The advent of artificial intelligence (AI) protein structure prediction programs, such as AlphaFold2 [24] and RoseTTAFold [25], has revitalized the potential of atomic integrative structural modeling. These programs can model the atomic structure of proteins and small complexes at a precision comparable to experimental structures. In some cases, AlphaFold structures have fit experimental data better than crystal structures [4] and may exhibit conformations resembling the complexed structure. Almost all proteins cataloged in UniProt are available for download from the EBI's AlphaFold database. Together with user-friendly implementations of these programs [26], we can predict atomic structures with relative ease.

Thus, the current state-of-the-art of integrative modeling is to build sub-complexes via AI-based structure prediction prior to assembling the higher-order structure. In our work on the human nuclear pore complex [4], we used AlphaFold to generate models of single proteins and sub-complexes that were assembled using our Assembline integrative modeling pipeline. The study exemplified several advantages of AlphaFold models for integrative modeling, such as an excellent fit to cryo-EM densities even if subunits are modeled outside the complex context, the surface complementarity of independently modeled proteins, and minimal steric clashes. Many studies have used AlphaFold models as starting structures for cryo-EM model building (for example: [27–30]), and some have begun using AlphaFold in cross-linking [31] and SAXS [32] studies. The structure of SMG1 kinase with its protein cofactors was determined by integrating cryo-EM and cross-linking data with AlphaFold model fitting and refinement [33]; cross-links were also used to evaluate the quality of AlphaFold models [34]. A new Phenix procedure integrated AlphaFold modeling and iterative refinement with high-resolution cryo-EM maps [35]. Although computation costs are still an issue, we can, nevertheless, confidently discontinue approximating structure with mere "beads" in integrative modeling.

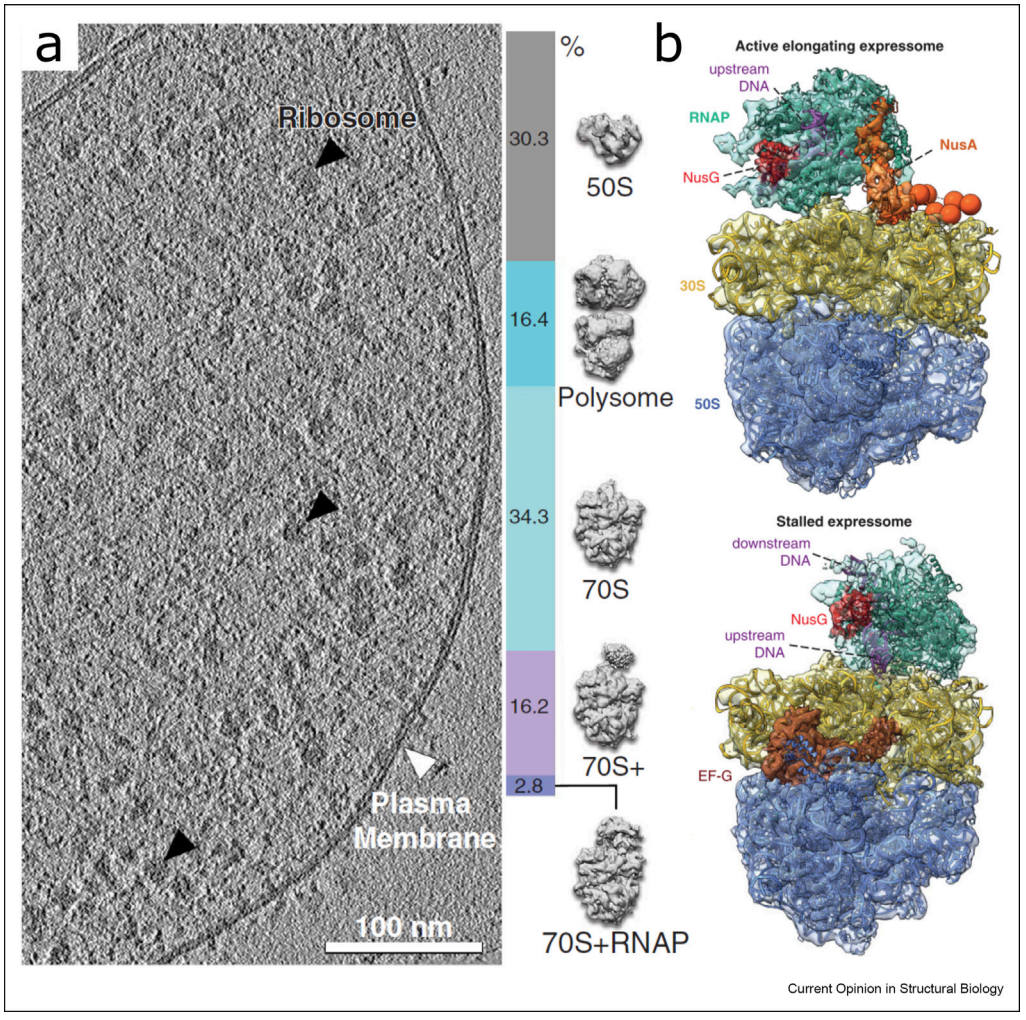
Can AI modeling programs reduce the number of orthogonal techniques needed for integrative modeling? On the one hand, it might be sufficient to obtain accurate models of sub-complexes with AlphaFold, put them together using a single experimental data set, such as cross-links, and only then validate against EM or SAXS data. This would simplify integrative modeling by eliminating the need for tedious restraint weight adjustments. On the other hand, the diversity of perspective is a key advantage of integrative modeling. A limitation of current AI-modeling tools is the nontrivial assessment of model accuracy [36]. Though they provide quality scores to estimate model accuracy, these are, nevertheless, just predictions with an inherent level of error. Modeling complexes is still challenged by unknown stoichiometry, alternative homo-oligomeric states (e.g., a tetramer vs. pentamer), and asymmetry. When AI-based models have wrong conformations or contain errors, additional experimental information will be required for refinement. For example, AlphaFold predicts an interaction between NUP160 and ELYS proteins in the human nuclear pore complex. However, the complex contains the NUP160 subunit on both sides of the nuclear membrane, but ELYS is present and interacts with NUP160 only on the nuclear side. In another example, XL-MS was used to guide the modeling of multi-domain SARS-CoV-2 proteins in conformations that could not be predicted using AlphaFold2 alone [37]. Thus, the biological context provided by techniques such as EM, cross-linking, SAXS, and biochemical experiments on interactions may still be necessary to clarify nuances and validate the

predicted models. Finally, the incorporation of additional data, such as EM densities, cross-link distance restraints, or SAXS profiles, into the AI-based structure prediction itself could create a new generation of integrative modeling programs, made possible by the availability of open-source codebases.

New opportunities in cellular structural biology

Cellular structural biology, whereby macromolecular structures are solved in the cellular environment, has gained traction thanks to recent advances in cryo-ET [38]. In this technique, cells are placed on a grid, vitrified, and, optionally, thinned into lamellae using focused ion beam (FIB) milling. Lamellae are imaged under a transmission electron microscope across a range of tilt angles, and the images are reconstructed into a 3D tomogram. Individual macromolecular particles can be computationally extracted, aligned, and averaged in a process called sub-tomogram averaging, resulting in 3D density maps with a resolution of 10—50 Å. FIB-milling is now an automated process and becoming faster and more reproducible [39], electron detectors are becoming more sensitive still, image acquisition is becoming more efficient [40], and processing software for finding and averaging particles in tomograms is becoming rapidly more streamlined and accurate [41,42]. With these advances in cellular cryo-ET, we can expect an explosive increase in cryo-ET output, akin to the 'resolution revolution' in single-particle cryo-EM [43]. Despite these advances and the first examples of high-resolution in-cell cryo-ET maps [41,42], cryo-ET

Figure 3



The structure of an actively transcribing-translating expressome determined by integrative modeling using in-cell cryo-ET and XL-MS data. (**a**) Slice through a tomogram of M. pneumoniae. Example ribosome particles are indicated by black arrows. Classification of particles on the right. (**b**) Models of the active elongating (top) and stalled (bottom) expressome. From O'Reilly FJ, Xue L, Graziadei A, Sinn L, Lenz S, Tegunov D, Blötz C, Singh N, Hagen WJH, Cramer P et al.: In-cell architecture of an actively transcribing-translating expressome. Science (80-) 2020, 369:554–557 [48]. Reprinted with permission from AAAS.

still suffers from comparably sparse data sets, resulting in low-resolution structures on average (10–50 Å). Cryo-ET is, therefore, in a prime position to benefit from integrative modeling. The most recent nuclear pore complex models are based on the in-cell cryo-ET maps [3,4,44–46]. Other examples include the Parkinson's disease-linked LRRK2 [47] and the complex of bacterial RNA polymerase in complex with a ribosome [48] (Figure 3).

A notable benefit of AI-based structure prediction for cellular cryo-ET structural biology is in providing structures for template matching—the identification of macromolecules in a tomogram based on a 3D search with a starting "template" structure [38]. Predicted models of complexes [49] could be used as better templates, enabling the identification of smaller and less abundant complexes, and a more complete interpretation of tomograms. Nonetheless, cryo-ET must still overcome many technical limitations. Even if 3D maps can be reconstructed, they may have an unknown identity and protein composition [50,51], unassigned densities [45], and unclear stoichiometry. Protein paralogs that exchange in the same complex in a context-dependent manner may not be distinguishable in 10–50 Å maps (e.g., different paralogs of tubulins in microtubules), leading to false positives in AlphaFold2 [36]. Thus, as with integrative structure determination of purified complexes, orthogonal in-cell techniques will be necessary for confident structure determination. These may include in-cell cross-linking [52], super-resolution microscopy [53], and cryo-super-resolution light and electron tomography [54]. Another opportunity lies in parallel single-particle cryo-EM analyses of cellular lysate fractions, containing dozens or hundreds of different complexes from the same cells that can be computationally identified and reconstructed, often at high resolution [55]. If high resolution cannot be obtained, XL-MS of the fractions can be used with integrative modeling, as was done for the 10-MDa eukaryotic pyruvate dehydrogenase complex [56]. Integrative modeling does not necessarily end when a structural ensemble is obtained. Rather, further data can be integrated into the model, such as biophysical simulations in the context of the native, cellular environment. For example, an integrative model of the nuclear pore complex was placed in the double membrane of the nuclear envelope to simulate pore dynamics [4].

Conclusions

Integrative structural modeling is an essential technique for structural biology. It combines the strengths of many techniques to resolve the structures of large and complex macromolecular systems which could not otherwise be achieved by a single experimental approach. Modern AI structure prediction tools have yielded structural models of proteins and small complexes at unprecedented precision, enabling us to overcome limitations in the scale and detail with which we can model macromolecular structures. To support the growth of integrative structural biology, we must also advance several key areas. It is imperative to integrate AI-based structure prediction tools with experimental restraints; develop a consistent rationale for automatic weighting of different data types; improve the computational efficiency of optimization algorithms; conform to accepted standards of integrative models, such as the inclusion of data and restraints in structural representations; and develop computational representations of novel in-cell data types. These developments will pave the way for ever larger and more detailed structural modeling studies, perhaps one day integrating with atomic biophysical simulations. Integrative structural modeling is well poised to solve problems in cellular structural biology, where high-precision data is more scarce, and multidisciplinary approaches are a necessity. We expect that future studies will broaden our perspective of integrative models beyond a structure in isolation and toward the cellular horizon.

Conflict of interest statement

Nothing declared.

Data availability

No data was used for the research described in the article.

Acknowledgements

We thank members of the Kosinski group and Grzegorz Chojnowski for feedback on the manuscript. Daniel S. Ziemianowicz has been supported by EMBL Interdisciplinary Postdoctoral Fellowship (EIPOD4) under Marie Sklodowska-Curie Actions Cofund (grant agreement number 847543).

References

Papers of particular interest, published within the period of review, have been highlighted as:

* of special interest
** of outstanding interest

1. Rout MP, Sali A: **Principles for integrative structural biology studies**. *Cell* 2019, **177**:1384–1403.

2. Alber F, Dokudovskaya S, Veenhoff LM, Zhang W, Kipper J, Devos D, Suprapto A, Karni-Schmidt O, Williams R, Chait BT, *et al.*: **The molecular architecture of the nuclear pore complex**. *Nature* 2007, **450**:695–701.

3. Akey CW, Singh D, Ouch C, Echeverria I, Nudelman I, Varberg JM, Yu Z, Fang F, Shi Y, Wang J, *et al.*: **Comprehensive structure and functional adaptations of the yeast nuclear pore complex**. *Cell* 2022, **185**:361–378. e25.

4. ** Mosalaganti S, Obarska-Kosinska A, Siggel M, Taniguchi R, Turoňová B, Zimmerli CE, Buczak K, Schmidt FH, Margiotta E, Mackmull M-T, *et al.*: **AI-based structure prediction empowers integrative structural analysis of human nuclear pores**. *Science* 2022, **376**, eabm9506.

This study used AlphaFold in combination with integrative modeling based on cryo-ET to build a structure of the human nuclear pore complex at near atomic precision.

5. Petrovic S, Samanta D, Perriches T, Bley CJ, Thierbach K, Brown B, Nie S, Mobbs GW, Stevens TA, Liu X, *et al.*: **Architecture of the linker-scaffold in the nuclear pore**. *Science* 2022:376. 80.

6. Farrell DP, Anishchenko I, Shakeel S, Lauko A, Passmore LA, Baker D, DiMaio F: **Deep learning enables the atomic structure determination of the Fanconi Anemia core complex from cryoEM**. *IUCrJ* 2020, **7**:881–892.

7. * Yu Y, Li S, Ser Z, Sanyal T, Choi K, Wan B, Kuang H, Sali A, Kentsis A, Patel DJ, *et al.*: **Integrative analysis reveals unique structural and functional features of the Smc5/6 complex**. *Proc Natl Acad Sci U S A* 2021, **118**, 2026844118.

A representative recent example of state-of-the-art integrative modeling using Integrative Modeling Platform that depicts all stages of the modeling and analysis.

8. Beckham KSH, Ritter C, Chojnowski G, Ziemianowicz DS, Mullapudi E, Rettel M, Savitski MM, Mortensen SA, Kosinski J, Wilmanns M: **Structure of the mycobacterial ESX-5 type VII secretion system pore complex**. *Sci Adv* 2021, **7**, eabg9923.

9. Dagenais P, Desjardins G, Legault P: **An integrative NMR-SAXS approach for structural determination of large RNAs defines the substrate-free state of a trans-cleaving Neurospora Varkud Satellite ribozyme**. *Nucleic Acids Res* 2021, **49**: 11959–11973.

10. Moysa A, Steczkiewicz K, Niedzialek D, Hammerschmid D, Zhukova L, Sobott F, Dadlez M: **A model of full-length RAGE in complex with S100B**. *Structure* 2021, **29**:989–1002. e6.

11. * Braberg H, Echeverria I, Bohn S, Cimermancic P, Shiver A, Alexander R, Xu J, Shales M, Dronamraju R, Jiang S, *et al.*: **Genetic interaction mapping informs integrative structure determination of protein complexes**. *Science* 2020:370.

The authors devised a new type of modeling restraint derived from the similarity between profiles of epistatic phenotypes obtained by screening single point mutations in the interacting proteins against secondary mutations in other genes. Residues which mutations lead to similar profiles were restrained to be close in space during integrative modeling leading to models as accurate as those obtained with crosslinking.

12. Webb B, Viswanath S, Bonomi M, Pellarin R, Greenberg CH, Saltzberg D, Sali A: **Integrative structure modeling with the integrative modeling Platform**. *Protein Sci* 2018, **27**:245–258.

13. Dominguez C, Boelens R, Bonvin AMJJ: **HADDOCK: a protein-protein docking approach based on biochemical or biophysical information**. *J Am Chem Soc* 2003, **125**: 1731–1737.

14. Saltzberg D, Greenberg CH, Viswanath S, Chemmama I, Webb B, Pellarin R, Echeverria I, Sali A: **Modeling biological complexes using integrative modeling Platform**. *Methods Mol Biol* 2019, **2022**:353–377.

15. Rantos V, Karius K, Kosinski J: **Integrative structural modeling of macromolecular complexes using Assembline**. *Nat Protoc* 2022, **17**:152–176.

16. Karaca E, Rodrigues JPGLM, Graziadei A, Bonvin AMJJ, Carlomagno T: **M3: an integrative framework for structure determination of molecular machines**. *Nat Methods* 2017, **14**: 897–902.

17. Ziemianowicz DS, altzberg D, Pells T, Alex Crowder D, Schräder C, Hepburn M, Sali A, Schriemer DC, Improv: **A resource for cross-link-driven structure modeling that accommodates protein dynamics**. *Mol Cell Proteomics* 2021, **20**, 100139.

18. Roel-Touris J, Don CG, Honorato RR, Rodrigues JPGLM, Bonvin AMJJ: **Less is more: coarse-grained integrative modeling of large biomolecular assemblies with HADDOCK**. *J Chem Theor Comput* 2019, **15**:6358–6367.

19. Viswanath S, Sali A: **Optimizing model representation for integrative structure determination of macromolecular assemblies**. *Proc Natl Acad Sci U S A* 2019, **116**:540–545.

20. Sinnott M, Malhotra S, Madhusudhan MS, Thalassinos K, Topf M: **Combining information from crosslinks and monolinks in the modeling of protein structures**. *Structure* 2020, **28**:1061–1070. e3.

21. Bonomi M, Hanot S, Greenberg CH, Sali A, Nilges M, Vendruscolo M, Pellarin R: **Bayesian weighing of electron cryo-microscopy data for integrative structural modeling**. *Structure* 2019, **27**:175–188. e6.

22. Viswanath S, Chemmama IE, Cimermancic P, Sali A: **Assessing exhaustiveness of stochastic sampling for integrative modeling of macromolecular structures**. *Biophys J* 2017, **113**: 2344–2353.

23. Ullanat V, Kasukurthi N, Viswanath S: **PrISM: precision for integrative structural models**. *Bioinformatics* 2022, https://doi.org/10.1093/BIOINFORMATICS/BTAC400.

24. ** Jumper J, Evans R, Pritzel A, Green T, Figurnov M, Ronneberger O, Tunyasuvunakool K, Bates R, Žídek A, Potapenko A, *et al.*: **Highly accurate protein structure prediction with AlphaFold**. *Nature* 2021, **596**:583–589.

This study presents AlphaFold2 – AI-based method that can predict structures of proteins at a precision similar to experimental structures.

25. Baek M, DiMaio F, Anishchenko I, Dauparas J, Ovchinnikov S, Lee GR, Wang J, Cong Q, Kinch LN, Schaeffer RD, *et al.*: **Accurate prediction of protein structures and interactions using a three-track neural network**. *Science* 2021, **373**:871–876. 80.

26. Mirdita M, Schütze K, Moriwaki Y, Heo L, Ovchinnikov S, Steinegger M: **ColabFold: making protein folding accessible to all**. *Nat Methods* 2022, **19**:679–682.

27. Andreeva L, David L, Rawson S, Shen C, Pasricha T, Pelegrin P, Wu H: **NLRP3 cages revealed by full-length mouse NLRP3 structure control pathway activation**. *Cell* 2021, **184**: 6299–6312. e22.

28. Noddings CM, Wang RYR, Johnson JL, Agard DA: **Structure of Hsp90–p23–GR reveals the Hsp90 client-remodelling mechanism**. *Nat* 2021, **601**:465–469. 6017893 2021.

29. Jones ML, Baris Y, Taylor MRG, Yeeles JTP: **Structure of a human replisome shows the organisation and interactions of a DNA replication machine**. *EMBO J* 2021, **40**, e108819.

30. Fontana P, Dong Y, Pi X, Tong AB, Hecksel CW, Wang L, Fu T-M, Bustamante C, Wu H: **Structure of cytoplasmic ring of nuclear pore complex by integrative cryo-EM and AlphaFold**. *Science* 2022, **376**, eabm9326.

31. Dawson JE, Smith IN, Martin W, Khan K, Cheng F, Eng C: **Shape shifting: the multiple conformational substates of the PTEN N-terminal PIP2-binding domain**. *Protein Sci* 2022, **31**, e4308.

32. Weber K, Doellinger J, Jeffries CM, Wilharm G: **Recombinant AcnB, NrdR and RibD of Acinetobacter baumannii and their potential interaction with DNA adenine methyltransferase AamA**. *Protein Expr Purif* 2022, **199**, 106134.

33. Langer LM, Bonneau F, Gat Y, Conti E: **Cryo-em reconstructions of inhibitor-bound smg1 kinase reveal an autoinhibitory state dependent on smg8**. *Elife* 2021:10.

34. Burke DF, Bryant P, Barrio-Hernandez I, Memon D, Pozzati G, Shenoy A, Zhu W, Dunham AS, Albanese P, Keller A, *et al.*: **Towards a structurally resolved human protein interaction network**. *bioRxiv* 2021, https://doi.org/10.1101/2021.11.08.467664.

35. * Terwilliger TC, Poon BK, Afonine PV, Schlicksup CJ, Croll TI, Millán C, Richardson JS, Read RJ, Adams PD: **Improved AlphaFold modeling with implicit experimental information**. *bioRxiv* 2022, https://doi.org/10.1101/2022.01.07.475350.

Presents a method for modeling based on high-resolution cryo-EM map that iterates EM refinement with AlphaFold modeling; at each iteration feeding the EM-refined model into AlphaFold as a new template.

36. * Allison TM, Degiacomi MT, Marklund EG, Jovine L, Elofsson A, Benesch JLP, Landreh M: **Complementing machine learning-based structure predictions with native mass spectrometry**. *Protein Sci* 2022, **31**:e4333.

A set of informative examples demonstrating various limitations of AlphaFold and the need of validating the models with orthogonal data.

37. Slavin M, Zamel J, Zohar K, Eliyahu T, Braitbard M, Brielle E, Baraz L, Stolovich-Rain M, Friedman A, Wolf DG, *et al.*: **Targeted in situ cross-linking mass spectrometry and integrative modeling reveal the architectures of three proteins from SARS-CoV-2**. *Proc Natl Acad Sci U S A* 2021, **118**:1–9.

38. Turk M, Baumeister W: **The promise and the challenges of cryo-electron tomography**. *FEBS Lett* 2020, **594**:3243–3261.

39. Berger C, Dumoux M, Glen T, Yee NB -y, Mitchels JM, Patáková Z, Naismith JH, Grange M: **Plasma FIB milling for the determination of structures in situ**. *bioRxiv* 2022, https://doi.org/10.1101/2022.08.01.502333.

40. Eisenstein F, Yanagisawa H, Kashihara H, Kikkawa M, Tsukita S, Danev R: **Parallel cryo electron tomography on in situ lamellae**. *bioRxiv* 2022.

41. Zheng S, Wolff G, Greenan G, Chen Z, Faas FGA, Bárcena M, Koster AJ, Cheng Y, Agard DA: **AreTomo: an integrated software package for automated marker-free, motion-corrected cryo-electron tomographic alignment and reconstruction**. *J Struct Biol X* 2022, **6**, 100068.

42. * Tegunov D, Xue L, Dienemann C, Cramer P, Mahamid J: **Multi-particle cryo-EM refinement with M visualizes ribosome-antibiotic complex at 3.5 Å in cells**. *Nat Methods* 2021, **18**:186–193.
The authors present a new method for generating 3D density maps from cryo-ET data that allowed them to obtain a density map at a resolution so-far unprecedented in cellular cryo-ET.

43. * Russo CJ, Dickerson JL, Naydenova K: **Cryomicroscopy in situ: what is the smallest molecule that can be directly identified without labels in a cell?** *Faraday Discuss* 2022, https://doi.org/10.1039/d2fd00076h.
This review presents the current state-of-the-art in celllar cryo-ET and discusses future directions.

44. Schuller AP, Wojtynek M, Mankus D, Tatli M, Kronenberg-Tenga R, Regmi SG, Dip PV, Lytton-Jean AKR, Brignole EJ, Dasso M, *et al.*: **The cellular environment shapes the nuclear pore complex architecture**. *Nat* 2021, **2021**, https://doi.org/10.1038/s41586-021-03985-3.

45. * Zimmerli CE, Allegretti M, Rantos V, Goetz SK, Obarska-Kosinska A, Zagoriy I, Halavatyi A, Hummer G, Mahamid J, Kosinski J, *et al.*: **Nuclear pores dilate and constrict in cellulo**. *Science* 2021, **374**, eabd9776.
This study uses integrative modeling based on in-cell cryo-ET data to demonstrate structural changes of the nuclear pore complex in cells.

46. Allegretti M, Zimmerli CE, Rantos V, Wilfling F, Ronchi P, Fung HKH, Lee CW, Hagen W, Turoňová B, Karius K, *et al.*: **In-cell architecture of the nuclear pore and snapshots of its turnover**. *Nature* 2020, **586**:796–800.

47. * Watanabe R, Buschauer R, Böhning J, Audagnotto M, Lasker K, Lu TW, Boassa D, Taylor S, Villa E: **The in situ structure of Parkinson's disease-linked LRRK2**. *Cell* 2020, **182**:1508–1518. e16.
A representative study demonstrating how automated integrative modeling can be applied to interpret low-resolution in-situ cryo-ET.

48. * * O'Reilly FJ, Xue L, Graziadei A, Sinn L, Lenz S, Tegunov D, Blötz C, Singh N, Hagen WJH, Cramer P, *et al.*: **In-cell architecture of an actively transcribing-translating expressome**. *Science* 2020, **369**:554–557. 80.
The authors used integrative modeling to determine a structure of a prokaryotic RNA polymerase in complex with the ribosome based on in-cell cryo-ET and crosslinking data.

49. Humphreys I, Pei J, Baek M, Krishnakumar A, Anishchenko I, Ovchinnikov S, Zhang J, Ness TJ, Banjade S, Bagde SR, *et al.*: **Computed structures of core eukaryotic protein complexes**. *Science* 2021:374. 80.

50. Klein S, Wimmer BH, Winter SL, Kolovou A, Laketa V, Chlanda P: **Post-correlation on-lamella cryo-CLEM reveals the membrane architecture of lamellar bodies**. *Commun Biol* 2021:1–12. 41 2021.

51. Martinez-Sanchez A, Kochovski Z, Laugks U, Meyer zum Alten Borgloh J, Chakraborty S, Pfeffer S, Baumeister W, Lučić V: **Template-free detection and classification of membrane-bound complexes in cryo-electron tomograms**. *Nat Methods* 2020, **17**.

52. Graziadei A, Rappsilber J: **Leveraging crosslinking mass spectrometry in structural and cell biology**. *Structure* 2022, **30**:37–54.

53. Liu S, Hoess P, Ries J. *Super-resolution microscopy for structural cell biology*, **51**; 2022:301–326. https://doi.org/101146/annurev-biophys-102521-112912.

54. Dahlberg PD, Moerner WE: **Cryogenic super-resolution fluorescence and electron microscopy correlated at the nanoscale**. *Annu Rev Phys Chem* 2021, **72**:253–278.

55. Skalidis I, Kyrilis FL, Tüting C, Hamdi F, Chojnowski G, Kastritis PL: **Cryo-EM and artificial intelligence visualize endogenous protein community members**. *Structure* 2022, **30**:575–589. e6.

56. * Kyrilis FL, Semchonok DA, Skalidis I, Tüting C, Hamdi F, O'Reilly FJ, Rappsilber J, Kastritis PL: **Integrative structure of a 10-megadalton eukaryotic pyruvate dehydrogenase complex from native cell extracts**. *Cell Rep* 2021:34.
Instead of purifying a complex or imaging it directly in cells, the authors take an intermediate approach analyzing a cell extract, which allows them to build an extended model of pyruvate dehydrogenase complex based on EM and crosslinking data.